Preetham Reddy Challa
AWS Certified Data Engineer

Python | SQL | AWS | Snowflakes | Tableau | Azure | Microsoft BI Stack (SSIS, SSRS, SSAS)

Phone: 3145391428
Email: preethamreddy1503@gmail.com
LinkedIn: https://www.linkedin.com/in/preetham-reddy-challa/

Certifications:

AWS Data Engineer Associate

Objective

Highly motivated Data Engineer/ Data Analyst with 4+ years of experience in designing, building, and optimising data pipelines, cloud infrastructure and machine learning models. Proficient in SQL, Python, and Bigdata technologies, with a strong ability to transform complex data into actionable insights.capable of leveraging AWS, Python and SQL for data engineering, visualization and cloud deployment.

Profile Summary

- Results-driven Data Engineer with a Master's degree in Computer Science and over 4 years of specialized experience in Data Engineering, Big Data, and Cloud Engineering, focusing on the Hadoop Ecosystem and Data Warehousing.
- Expert in designing and optimizing data pipelines using Python, SQL, and big data technologies like Hadoop, Apache Spark, and AWS services (EC2, S3, EMR, RDS, Lambda), ensuring efficient data processing and integration.
- Skilled in cloud infrastructure management across AWS and Azure platforms, with hands-on experience in deploying scalable solutions using Kubernetes, CloudFormation, and PowerShell.
- Proficient in ETL tools like Informatica PowerCenter and SSIS, streamlining data processes and enhancing accessibility, while supporting complex analytics across major Hadoop distributions like Cloudera and Azure HDInsight.
- Advanced data visualization and analysis using Tableau, Power BI, and Python libraries (Pandas, Matplotlib), delivering actionable insights through interactive dashboards and reports.
- Strong collaborator with cross-functional teams, adept in Agile and Waterfall methodologies, consistently ensuring timely project delivery and supporting data-driven decision-making across the organization.

Technical Expertise

Programming and Scripting: Python, SQL, Java, R, 4GL

Cloud Platforms and Services:AWS (S3, EC2, lambda, Redshift, Glue), Azure(Azure Networking, Azure Storage, Azure virtual Machine)

Databases: Mysql, MongoDB, Oracle, PostgreSQL, Snowflake

Big Data Technologies: Hadoop, Spark, Hive

Data Storage & warehousing: Snowflake, Amazon Redshift

Data Visualization: Tableau, Power BI, MS Excel

ETL/ELT Tools: Informatica, PowerCenter, Talend, Apache Airflow

Project Management Methodologies: Agile(Scrum), Waterfall

Operating Systems: Windows, Linux, Unix

Version Control: Git, GitHub

Tools and Others: Docker, Pycharm, Jupyter Notebook, Infor ERP Tools

Professional Experience

United Health Care **MO,USA**

Data Analyst/Data Engineer **(Jan 2024 - Present)**

- Independently designed scalable data pipelines using Python, processing over 500,000 records from various sources monthly, reducing operational costs.
- Administered Epic system components, managed user accounts/roles, implemented security permissions, and continuously monitored system health and performance to ensure optimal operation and HIPAA compliance.
- Implemented a comprehensive Lambda processing flow for unstructured data, including metadata separation and efficient handling of missing files through a Dead Letter Queue (DLQ) mechanism.
- Contributed to the migration of Epic Systems to Azure, assisting with Azure infrastructure setup (VMs, Storage, Networking), executing data migration plans, and performing thorough testing and validation to ensure system integrity and performance post-migration. Provided go-live support for a smooth transition.
- Managed Epic system updates and patches, applying necessary enhancements and security updates while providing training and technical support to end-users to improve system utilization
- Created interactive dashboards (Tableau, Excel) translating complex data (potentially specified if Epic/Healthcare data) into insights for stakeholders, reducing project delivery time by 5 days.
- Optimized PostgreSQL database performance through advanced query tuning, strategic indexing, and efficient schema design. This led to a reduction in data retrieval times by 50 milliseconds per query, supporting real-time analytics and contributing to a client retention increase.
- Conducted root-cause analysis of data anomalies in healthcare provider networks, leading to a 20% reduction in claim discrepancies.
- Developed big data solutions utilizing Apache Spark, HBase, and AWS EMR, enabling distributed processing of datasets exceeding 1 million rows. The solution achieved query response times of under 2 seconds, enhancing business operation productivity and providing timely insights for strategic decisions.

- Implemented and managed AWS cloud infrastructure including S3, EC2, and RDS, which supported scalable data storage and processing solutions. This resulted in a cost saving and ensured high availability and reliability of critical business applications.
- Collaborated on the development of 15 end-to-end ETL processes using SSIS and SSRS, automating data extraction, transformation, and loading procedures. This automation reduced manual data handling by 70 hours per month and decreased error rates by 20%.
- Utilized Pandas for advanced data manipulation and analysis, cleaning and transforming large datasets efficiently. This improved the accuracy and reliability of data used in downstream processes, leading to a reduction in reporting errors.
- Developed and optimized complex Excel models for detailed financial and operational analysis, enabling the business to simulate 5 distinct scenarios. This provided key insights that contributed to a revenue increase from optimized resource allocation and strategic planning.
- Worked closely with cross-functional teams, including data scientists, project managers, and business analysts, to ensure technical solutions aligned with business objectives. This collaboration resulted in improved project delivery timelines by 10 days and enhanced team communication and problem-solving.

Tools & Technologies: SQL, Scala, kafka, Snowflakes, Python, Dynamodb, Git, Docker, Tableau.

Infor **Hyderabad, India**
Data Engineer/Project Engineer **(May 2021 - Dec 2022)**

- Independently built ETL pipelines using Python and SQL for streamlined data integration across multiple departments.
- Managed relational databases in Oracle, implementing advanced indexing techniques, optimized query structures, and partitioning methods. These improvements ensured data consistency and enhanced system performance, reducing query execution time by 200 milliseconds and increasing overall system resilience.
- Created over 50 visual dashboards using Matplotlib, providing stakeholders with clear and actionable insights into data flow, processing, and integration. These visual tools improved stakeholder oversight and contributed to a 20% increase in project transparency and decision-making efficiency.
- Monitored data integration tasks in real-time using AWS CloudWatch and automated large-scale ETL processes with AWS Glue. These efforts significantly improved platform stability, reducing downtime by 30%, and ensured that the architecture was scalable for future data integration needs.
- Processed and ingested raw data into the platform using Hadoop and Pig, ensuring the platform's capacity to handle large data streams necessary for research and analytics. This capability supported the analysis of datasets exceeding 5 million records, enabling complex analytics and improving research outcomes.
- Developed and delivered 25 dashboards and 40 reports using Power BI and MS Excel, making data integration performance, data quality, and platform efficiency accessible to over 100 users across 5 departments. This work facilitated data-driven decisions that contributed to increase in operational efficiency.

- Configured and optimized Informatica Power Center and SSIS environments for data integration, enhancing data accessibility and enabling complex analytics and reporting initiatives. This optimization supported over 10 key analytics projects, improving data accuracy and availability.
- Applied the Waterfall methodology throughout the platform's development, ensuring on-time delivery of a high-quality data integration solution. The project was completed within a tight deadline, meeting all client requirements and contributing to increase in client satisfaction scores.
- Documented over 20 changes to ETL processes, data models, and integration scripts using GitHub. This meticulous documentation fostered collaboration across 3 development teams, ensuring traceability and accountability, which improved project efficiency.
- Developed, tested, and documented data processing scripts in Jupyter Notebook, ensuring accurate and efficient data transformations. This contributed to a 10% reduction in errors during data integration and improved the reliability of the data processing pipeline.

Tools & Technologies: GCP, SQL, Epic Systems, Scala, Big Query, Kafka, Data Studio, Snowflakes, Looker, Azure, Machine Learning, Tableau.

Infor **Hyderabad, India**

Associate Consultant: **(Jan 2021 - May 2021)**

- Provided Tailored ERP (Enterprise Resource Planning) solutions to clients according to their requirements within the platform.
- Transformed the huge amount of data from on premises to on cloud ensuring the security and high performance.
- Developed and optimized reports and data warehouse components for financial and operational analysis.
- Developed the new warehouse and report components for the client for the migrated data.
-

Tools & Technologies: Infor Ln, 4GL, Infor ERP.

Projects

- **Real-Time Data Pipeline for Event Monitoring and Analytics**
    - Built a real-time data pipeline to monitor and analyze events like user activity, transactions, or system logs.
    - Implemented data processing using Spark Streaming.
    - Designed a schema to store the processed data in a data warehouse optimized for analytics.
    - Created dashboards to monitor key metrics such as user activity trends or error rate spikes.

  Tools & Technologies: Apache Kafka, Apache Spark,Snowflake, Amazon Redshift, Tableau.

- **Netflix Dashboard – Tableau Project**
    - Built an interactive dashboard using Tableau to analyze Netflix data, including popular genres, viewer ratings, and global reach.
    - Visualized complex datasets through dynamic charts, graphs, and heatmaps to highlight key trends.

- Focused on usability by incorporating filtering and sorting features for tailored insights.
- Demonstrated strong data cleaning, transformation, and visualization skills.

Tools & Technologies: Tableau, SQL, Excel, Python, Data Visualization

Achievements

- Certified in “Programming Data Structures and Algorithms Using Python” in Swayam NPTEL with 76%.
- Certified by CCNA in Basic of Networking and Cisco Packet Tracer.

Education

Master of Science in Computer Science **(Jan 2023-May 2024)**

University of Central Missouri

Relevant Courses: Advanced Algorithms, Advanced Database Systems, Prog Found Data Science AI.